


```
Db 70 QHRHCO-CGAARPAALUSSPKPTVTQTTTAPGAPGFSHWSQPOLHAAVLRAVLADIGEVDITA 128
QY 72 GDC-----HLTEEEFGLV-----QSMKCD--TVRIKGVLOQPTT 105
Db 129 GACCTARGTTRTATQAQLPMSHLGVSVCNGECVPAAGSSGGCGACTCRGCGCLDGP-- 186
QY 106 APPLMTSEGNVTAEDTEFAIRAFVYAAASAAEAHWHRLVLLSQIHEPIGSGGNIIN 165
Db 187 -----NASRDEQOCRN-----HRRH-----RSSHRKRHRGSSRRR 217
QY 166 TNKGGRSCONPALPSPDQSPSGNATTSVTRDNYHLLTEEEFGVWSQMKHSONKSGGSV 225
Db 218 RDGSAGSCP-PSPGSGSKPT--ATTARLPSAVSLLA-----TASNAS 257
QY 226 PVRGPTQPCSESOILKSEFVPTPKENKQER-----EDENW----- 264
Db 258 PVASPRDOEVLIRIMAKYRTPPTPTESAQQOPLHIFLHCECEKTAAPGVYALTGTSA 317
QY 265 -----RLPPPPVAETVPSPSVTEIETPLQRIPTATIAEPLGH 304
Db 318 SAAAAVAPGLAAKATATASPSPPPPAASAPPPPPAAS-VPPP-----PPAASVPPPPP 369
QY 305 CFTTISPAFVHSLNKRKQLELLREVEWPGRGHMAATCCCKLOVEGQDRTHSL-----A 359
Db 370 -----PA-----VSVPPPPRAMSIPLPPPA 389
QY 360 AA-----PVREAPPPPTGASPSVPALPGADPQPSAELL 394
Db 390 ASIPPLPAVSPPPPPAASVPPPLPAVSPPPPPRSASAL 429

RESULT 2
Q9NEEO PRELIMINARY; PRT; 1246 AA.
AC Q9NEEO;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 129.9 kDa protein.
GN p1105.03
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL139794; CAB75561.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1246 AA; 129927 MW; AA5AFF03EE0D7B0 CRC64;

Query Match 6.2%; Score 135; DB 5; Length 1246;
Best Local Similarity 20.1%; Pred. No. 0.061;
Matches 104; Conservative 54; Mismatches 158; Indels 202; Gaps 24;

QY 21 QASVLSLITENOGKRCPCGAQNLMTQNPTLPVS---SHRSPGNAASVVTGGDCHLP 77
Db 195 QTSVSTPP-----RVP-----QACANAKSASVVDGSRASATTLSHPDVAVIA 240
QY 78 TEEFGLVQSMKCDTVRIKGVLOQPTTAPPLMTSEGNVTAEDTEFAIRFV--YAVAAA 135
Db 241 SD-----VSLSC-----ASAAPMSIERDDSDANCIDENEPEFCVDYTWSEAA 282
```

```
QY 136 SAAEAHWHRLVLLSQIHEPIGSGGNIINTNKGGRS-----CONPALPSPDQSPSGN 188
Db 283 SAAPALH-----RHTPSTVAAVTRDEKGRNGSCRODTGCGSPPLVHSAASRGDGE 332
QY 189 ATTSVTRDNYHLLTTEE-----FGVWSQSMKWHSONKSGSVVRCPTQPCSESOILKE 243
Db 333 GCGAVDDQEAHTLAQOEKAAESCGAW-----ESATATTTTPVR--FYRLYCR--YRT 379
QY 244 SEVPPTTPKENKQRED-----ENWRLPPPPVA--- 272
Db 380 SHDTSPSAGGNGNEEDEAEHCGRGEEDSNVGFSLATSPSAAAEQWR--PCPVVGTD 437
QY 273 ---ETPVSPSVTEIET---PLQRIPTATIAE-----PLGH-----CTFTISP-- 311
Db 438 VGHSTPSPKPTVVLEDTMTALAVSALPCGLTRNGEDSGRLDSPMRHRTARVSTMSPVV 497
QY 312 ---AFVHSLNKRKQLELLREVEWPGRGHMAATCCCKLOVEGQDRTHS-----LAAAPVR 364
Db 498 DPNGLVFSPSCRQMRRLQRR-----CPRQTSEGSDAEYAPHPTVPVGPSPQ 544
QY 365 EAPPPPT-----GASPSVPALP----- 383
Db 545 SPRPPPSCHWSTPLLRRCRRDAQTPAPAPLSAASQPPVPSPAPSPDTMDMADRGAGATDEL 604
QY 384 -----GADPQR-----SAELLLLAVTREGLERR 406
Db 605 AGSHAADAAMGASPSRVIHLKKEKAVTVTSNIGQERR 642

RESULT 3
Q96Q53 PRELIMINARY; PRT; 562 AA.
AC Q96Q53;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Aristalless-related homeobox protein ARX (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohira R.H., Zhang Y.H., Guo W., Dipple K., Shih S., Doerr J.,
RA Huang B.-L., Fu L., Abu-Khalil A., Geschwind D., McCabe E.;
RT "Human ARX gene: genomic characterization and expression.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY038071; AAK93901.1; -.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR Pfam; PF00046; homeobox_1.
DR ProDom; PD000010; Homeobox_1.
DR PROSITE; PS00027; HOMEBOX_1; UNKNOWN_1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 562
SQ SEQUENCE 562 AA; 58159 MW; FBDF41E387C65532 CRC64;

Query Match 6.1%; Score 134.5; DB 4; Length 562;
Best Local Similarity 20.5%; Pred. No. 0.025;
Matches 101; Conservative 33; Mismatches 143; Indels 215; Gaps 18;

QY 15 QLGQALQASVLSLITENOGKRCPCGAQNLMTQNPTLPVSVSHRSPGNAASVVTGGDC 74
Db 160 KISQAPQVSISSKSYRENG--APFV-----PPPPALDELGGP-----GGVT 199
QY 75 HLPTEEEFGLVQSMKCDTVRIKGVLOQPTTAPPLMTSEGNVTAEDTEFAIR----- 126
Db 200 H-PEER-----LVAGGPGSAP--AAGGTCTDEDEELEDDEDEE 240
QY 127 -----AFVYAVAAASAAEAHWHRLVLLSQIHEPIG 158
Db 241 EELLEDEELEDDEDDARALLKEPRRCPVAAATGAVAAAAA----- 280
```



```
QY 2 GWSVVVLCGKQALQASVSLSIITENOGKRCPCFGAQNLMTCQNTLPVSHSRP 61
D 188 GSYTQTVPTGDKALSERFGSGAATETAEARRSSSDTAAY-----P 230
QY 62 PCNAASVTGGDCHLPTEEEFGLVQSMKCDTVRKGLVQGTTPATPLMTSEGNV----- 116
D 231 AGTTAVTGGNGTQCSQDTSE-----SSRQDTPSSFGQFTPOSSQGTPTYSKSTGTPYSQD 286
QY 117 TAEDTEEIRAFAVAVAAASAAEAHWRHLVLLSQIHEPIGSGGNIINTNKGGRSCONP 176
D 287 SAYSSSTTSTFKPRSENSYQDAFRRHFSASSASTTASTAIAATTAATASSASSSSL 346
QY 177 ALPSPDQSPGNATVTRDNYHLLTEEEFGVWSQSMKWHQSGGVVVRGPTQEPSCS 236
D 347 SSSSSSSSSSSQSPRSSDANYPAYES-----WNRYSR-HT-----SYPPRRATREE-- 393
QY 237 ESQILKESFVPTTPKNNKQOERDENWR--LPP-----PPVAETPV----- 276
D 394 -----PPGAPFAENTAERFPPTSLSYLPPEPSRPTDQYRPPASEAPPPPPPPFG 443
QY 277 -----PSPSVTELETPLQIPRTATIA--GEPLGHCTFTTSPAFVHVLNKKRQLEL 327
D 444 GGGGGGSPREEVYRTS-----PRPASRGSPAPETNVSVPQAQHSLLDSR---TEM 496
QY 328 LLRE-----VEMP-GRGH----- 339
D 497 LLKQSRKFSFLASDTEEEENSSVMVLGARDTGSSEVSGSGHGPCTPPAPANFEDVAPT 556
QY 340 -----MAATCK---LOVEQDRTWSLAAAPV---REAPPPPTGASS 375
D 557 GSGEPGATRESPKANGQNAQSCSGDDMETSDDRGSGPPAPTTPQOQPPPPPPPPPP 616
QY 376 EPSVPALP-GADPORSALL 394
D 617 PPLASLPLGYPHOPAYILL 636

RESULT 12
Q9VZU6
ID Q9VZU6 PRELIMINARY; PRT; 828 AA.
AC Q9VZU6; 024005;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BTBVII protein (BTB-VII protein domain).
GN BTBVII OR CG11494.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
```

```
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RN SEQUENCE OF 4-118 FROM N.A.
RX MEDLINE=95024186; PubMed=7938017;
RT Zollman S., Gott D., Prive G.G., Couderc J.L., Laski F.A.;
RT "The BTB domain, found primarily in zinc finger proteins, defines an
RT evolutionarily conserved family that includes several developmentally
RT regulated genes in Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:10717-10721(1994).
DR EMBL; AE003477; AAF47721.1; -.
DR EMBL; U14404; AAA50839.1; -.
DR FlyBase; FBgn0012049; BtbvII.
DR InterPro; IPR000210; Btb_P0Z.
DR Pfam; PF00651; Btb; 1.
DR SMART; SM00225; Btb; 1.
DR PROSITE; PS00097; Btb; 1.
SQ SEQUENCE 828 AA; 87135 MW; 19A72E077D570B41 CRC64;

Query Match 5.9%; Score 128.5; DB 5; Length 828;
Best Local Similarity 22.3%; Pred. No. 0.12;
Matches 100; Conservative 44; Mismatches 134; Indels 171; Gaps 22;

QY 14 KOLGALQASV-----SLSIITENOGKRCPCFGAQNLMTCQNTLPVSHSRPPGNAAYS 68
D 488 QQQQQQEQEASASQAYSQIITVNNLVGSYATAAQNLT-----SPTSPNESN----- 533
QY 69 VTGGDCHLPTEEEFGLVQSMKCDTVRKGLVQGTTPATPLMTSEGNVTAEDTEEIRAFA 128
D 534 -----MQQSV-----YSQGT-----PTQSPVHA- 552
QY 129 VYAVAAASAAEAHWRHLVLLSQIHEPIGSGGNIINTNKGGRSCONPALP-SPDQSPSG 187
D 553 --GVGASAG-----GAAGNASAGNGAGGAQANQVVKRSVNPQG 592
QY 188 NATTSVTRDNYHLLTE-----EEFGVWSQSMKWHQSGGVVVRGPTQBP 234
D 593 D-----ENFTRALEAVRTGGIGFCKAARLYGVNVRTL-WLEYKKRG-----YPVSRP 638
QY 235 CSESQILKE--SFVPTTPKNNKQOERDENWRLLPPPPVAETPVSPSVTELETPLQIRP 292
D 639 SIKARVWQEPNLSPSTPSTNGDDNTNETLGMQIPPOAETPTPSLMCTSHHTGL----- 694
QY 293 RTATIAGE-PLG---HCTFTI-----SPAFVHVLNKKRQLELLREVWEPGRG 338
D 695 --GSAGSLPAGGNHHPALGVMSLFDPRYMSDFGNVHSMTRQ-----RYIATGGG 743
QY 339 HMAATCC-----KLQVEQDRTWSLAAAPVREAPPPTGASSEP-----SYVALPGADP 387
D 744 AGAGTGAGTGTTNNTISSQAGTATLQAAAVN-----AASEPAESLSISAMPVSPV 794
QY 388 QRSAEILLL-----AVTREGLE 404
D 795 MGAHSFLINPVTVAAPPVATVTSTGLQ 823
```



```
DR EMBL; AB026671; BAA85852.1; JOINED.
DR EMBL; AB026672; BAA85852.1; JOINED.
DR EMBL; AB026673; BAA85852.1; JOINED.
DR HSP; P06601; IRLJ.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 562 AA; 58215 MW; A64AELC351FC302 CRC64;

Query Match          5.8%; Score 127; DB 11; Length 562;
Best Local Similarity 20.4%; Pred. No. 0.095;
Matches 103; Conservative 33; Mismatches 130; Indels 238; Gaps 20;

QY 15 QLQALQASVSLSTITENQGRKPCGCAQNLMTCNFTLPVSHRSPGNAASVVTGGDC 74
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 159 KISQAPQVSISRKSYRENG--APFV-----PPPPALDELSGPGVA----- 198

QY 75 HLPTEEEGVLVQSMKCDTVRIKVGLOGPTTAPPLMTSEGNVTAEDTEEA----- 125
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 199 H-PEER-----LSAASGPGSAP---AAGGTGAEDDEEELLEDEDEE 238

QY 126 -----RAFV-----YAVAAASAAEAHWRHLVLLSGQIHE 155
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 239 EEELEDDDELEDDARALLKEPRCSVATTGTVAASAAAAA----- 283

QY 156 PIGSGGNIINIKGRSCONPALPSPQSPSGNATTSV----- 193
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 284 -----VATEGGELSPKEELLHPEDAEGKGDGDSVCLAAGSDSEGLLKRQRYRT 335

QY 194 TRDNYHL-----LTEEEFGVWSQS--MKWHSQNKSGG--- 223
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 336 TFTSYQLEELERAFQKTHYPDVFTTEELAMRLDLTEARVQVWFQNRRAKRRKAGAGQTH 395

QY 224 --SVPVRGPTQPCSESOILKESFVPTTPKENNKQEREDNW-----RLPPP 269
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 396 PPGLPFGPLSATPLSPYLDASFPFPHHP-----ALDSAWTAAAAAAPPSPPLPP 448

QY 270 PVAETPVPSVTEIETPLQRIPTATITAGEPLGHCTFT-----ISPAFVHVSYN 319
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 449 PGSASLPPS-----GAPLGLSTFLGAAVFRHPAFISPAF----- 482

QY 320 KRKQLELLREVEWPGRGHMAATCCKLQVEGQDRTWSLAAAPVREAPPPPTGASSEPSV 379
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 483 -----GLFSTMAPL-----TSASTAALLRQPTPAVEGAVASGAL 517

QY 380 --PALPGADPORSALLLLAVTRE 401
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 518 ADPATAANDRRASSIALRLKAKE 541
```

Search completed: December 25, 2002, 01:13:08
Job time : 89 secs

THIS PAGE BLANK (USPTO)